Analysis of factors affecting the birth of low weight babies

Nisanth Dheram, Sridevi Bayyapuneedi, Vaishnavi Jala

Executive Summary:

According to WHO, a baby is considered to be of low weight if it weighs less than 5 pounds 8 ounces at birth (2500 grams). Low baby birth weight is a major concern in United States because of its high mortality rate and a higher probability of leading to serious medical conditions. The goal of this analysis is to identify risk factors associated with the mother of giving birth to a low birth weight baby. We considered factors such as *age, weight, race, hypertension, uterine irritability, premature labor history*, *smoking status* and *number of physician visits.*

From the analysis, we observe that the mother's weight at her last menstrual period is positively correlated with baby birth weight. Any medical conditions like hypertension and uterine irritability have a negative impact on baby birth weight. Smoking is also found to have an impact on baby birth weight and is found to be lesser for babies of mothers who smoke. Further analysis showed that the mean baby birth weight is the least for African American mothers compared to those belonging to white and other races.

Finally, we conclude that the mother's behavior rather than her physical characteristics has a higher impact on the baby's birth weight.

Introduction:

The goal of this analysis was to identify risk factors associated with the mother of giving birth to a low birth weight baby. WHO states that a baby is considered to be low weight if it weighs less than 5 pounds 8 ounces at birth (2500 grams). Low birth weight in general is thought to place the infant at greater risk of later adult chronic medical conditions such as diabetes, hypertension, and heart disease. Low birth weight infants account for much of the increased morbidity and mortality. In 1985, low birth weight babies comprised only 6.8 percent of all live births but 61.1 percent of all infant deaths in United States. Given this, understanding the effects of the mother's characteristics and behavior on a newborn is vitally important to better-ensuring healthy pregnancies and births.

To analyze the birth weight of baby, a study was conducted at Baystate Medical Center, Springfield, Massachusetts during 1986. Data were collected on 189 mothers, 59 of whom had babies of low birth weight. We will be using this data to carry out our analysis. The data collected has the following variables:

Variables relating to the mother's demographics: *Age, Weight (at last menstrual period), Race*

Variables relating to the mother's medical conditions: *Hypertension, Uterine Irritability, Premature labor history*

Variables relating to the mother's behavior: *Smoking status, Number of physician visits*

Following is how each variable looks like in the data:

Predictor	Description
Age of Mother	Years
Weight at last menstrual period	Pounds
Race	White, African American, Other
Smoking Status	Yes/No
Premature labor history	Number of premature births
Hypertension history	Yes/No
Uterine Irritability	Yes/No
Physician visits	Number of visits

Analysis:

The preliminary analysis explains each factor's distribution and its relationship with baby birth weight. Next, regression is used to determine the factors which collectively explain baby birth weight. From this, we identify the significant factors and analyze the effect of them on baby birth weight individually. Finally, we conclude the most significant factors leading to low birth weight in babies and the precautions to be taken to minimize the risk of low baby birth weight.

Data Description:

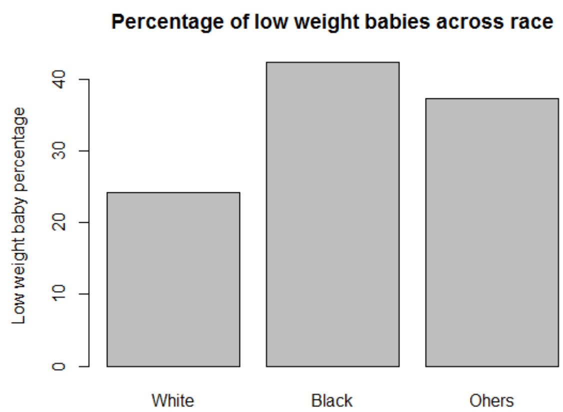
We observed the distribution of each factor to identify outliers and missing values. We then analyzed the relationship of each factor with the dependent variable, i.e., baby birth weight. The analysis of each factor is as follows:

Age of the mother:
The distribution of this variable is slightly skewed to the right. We observe that there is an outlier sample where age of the mother is 45. Since this sample deviates significantly from the rest of the observations, we have removed it for a better fit of our regression model. The mean age of the mother is 23.24 pounds and the median is 23 pounds.

Weight of the mother at last menstrual period:
The distribution of this variable is skewed to the right with a mean weight of 129.9 pounds and a median of 121 pounds. We observe that baby birth weight increases linearly with the mother's weight. (Refer to Appendix 1. a.)

Race:
This is a nominal variable which differentiates if a mother is white, African American or of some other race. 95 of the mothers are white, 26 are African American and the rest 67 belong to other races. The mean baby birth weight of white mothers is 6.8 pounds, followed by 6.2 pounds for those of other races and 6 pounds for mothers who are African American. The proportion of low birth weight babies is 42.3%

in African American mothers and white mothers have 24.2% of low birth weight babies. (Refer to Appendix 1.b.)



Smoking Status:
This factor is a nominal variable which identifies whether a mother smokes or not. 74 out of 188 mothers are found to smoke. The mean baby birth weight of mothers who smoke is 6.1 pounds which is less than 6.7 pounds for mothers who do not smoke. The proportion of low birth weight babies for mothers who smoke is 25.4%, which is higher compared to 40.5% for mothers who do not smoke.

Premature Labor History:
This factor represents the number of premature births a mother has had before. This variable is highly skewed to the right with the mean of 0.2 and median of 0 premature births. We observe that baby birth weight decreases with more premature births and the relationship is statistically significant. (Refer to Appendix 1.c.)

Hypertension:
This factor gives information about whether a mother has hypertension or not. 12 out of 188 of the mothers have hypertension. The mean baby birth weight of mothers with hypertension is 6 pounds which is less than 6.5 pounds for mothers who do not have this condition. The proportion of low birth weight babies for mothers with hypertension is 58.3%, which is higher compared to 41.7% for mothers without hypertension.

Uterine Irritability:
This factor represents if a mother has uterine irritability. 28 out of the 188 mothers have uterine irritability. The mean baby birth weight of mothers with uterine irritability is 5.4 pounds which is less than 6.6 pounds for mothers without uterine irritability. The proportion of low birth weight babies for mothers with uterine irritability is 50%, which is higher compared to 28.1% for mothers without this condition.

Physician Visits:
The distribution for the number of visits to a doctor during the first trimester is highly skewed to the right with a mean of 0.79 and a median of 0. We observe that there is no significant linear pattern

between number of physician visits and new born baby birth weight. The mean baby birth weight is minimum of 5.5 pounds for 4 physician visits and maximum of 7.3 pounds for 6 physician visits.

Baby birth weight:
Baby birth weight is normally distributed with mean of 6.5 pounds and median of 6.6 pounds. 31.3% percent of the mothers have low birth weight babies. The mean weight of low birth weight babies is 4.6 pounds compared to 7.3 pounds for others. (Refer to Appendix 1.d.)

Statistical Model Analysis:

We have used a regression model to identify the factors which collectively have a statically significant impact on baby birth weight. (Refer to Appendix 2 for a more detailed analysis for the selection the model). The model is found to explain 23.5% of the variation in baby birth weight. The model is represented by the equation:

Baby birth weight = 6.5 + 0.01*(Weight of mother)
- 0.73*(Smoking Status)
- 1.3*(Hypertension)
- 1.15*(Uterine Irritability)
- 1.1*(Race: African American)
- 0.75*(Race: Others)

From the above equation, we can interpret the following:

- When weight of mother decreases by 1 pound, baby birth weight decreases by 0.01 pounds
- The weight of a baby varies between mothers who smoke and who do not smoke. The baby birth weight decreases by an average of 0.73 pounds for mothers who smoke
- Women with hypertension have babies weighing 1.3 pounds less than those who don't
- Women with uterine irritability have baby birth weight of 1.15 pounds lesser than those who do not have this condition
- Baby birth weight varies with the race of the mother. Average baby birth weight of mothers belonging to African American race is 1.1 pounds lesser compared to white mothers and average baby birth weight of mothers of other races is 0.75 pounds lesser than white mothers

Note: The above interpretations only hold for mothers who have the same characteristics and behavior for variables other than the one being changed.

Conclusion:

Our final model describes the relationship between birth weight of babies and the mother's physical, behavioral and medical characteristics. Our analysis shows that the effect of weight of the mother at her last menstrual is significant to baby birth weight. We also observed that baby birth weight does not depend on age of the mother. The number of physician visits during the first trimester of pregnancy is not significant in describing low birth weight. Any medical conditions the mother has like hypertension and uterine irritability have been found to have a negative impact on baby birth weight. The mother's smoking habits also have a negative impact on birth weight of a baby.

Finally, to reduce the likelihood of low birth weight babies, a mother should be a non-smoker, free from hypertension, free from uterine irritability and with healthy weight at the last menstrual period.

Mothers of African American and other races need to give more importance to the above precautions compared to whites. Proper health care during pregnancy and management of high-risk factors may reduce the incidence of low birth weight babies.
Refer to Appendix 3 for assumptions made and collinearity checks performed.

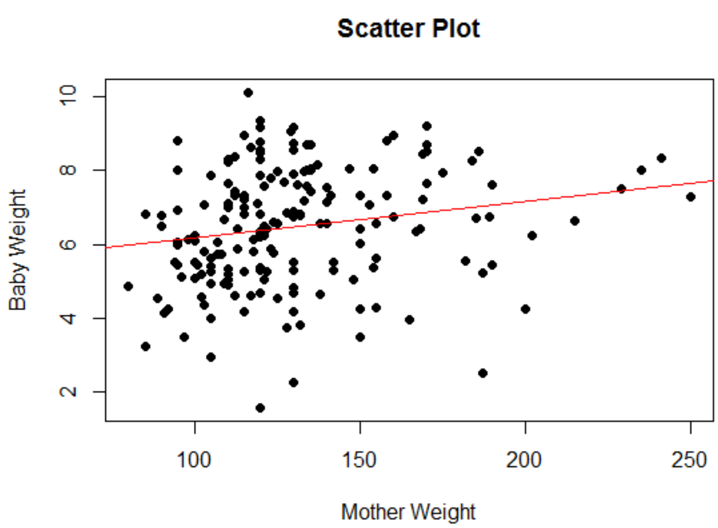
Limitations:

- The model is able to explain only 23.5% of the variation in baby birth weight. There might be other external factors as below which could account for the unexplained 76.5% variation
    - Food habits
    - Alcohol consumption
    - Family history
    - Multiple pregnancy (twins, triplets etc.…)
- The data collected in Springfield, Massachusetts, United States. This model might not be applicable in other countries like United Kingdom, China.
- The study conducted in 1986, so the factors in the analysis might not behave in the same way during the recent years.
- The range of mother’s age is between 14 and 36 years and mother’s weight is between 80 and 250 pounds. Baby birth weight of mothers whose physical characteristics are not within this range might not be explained using this model.

Appendix:

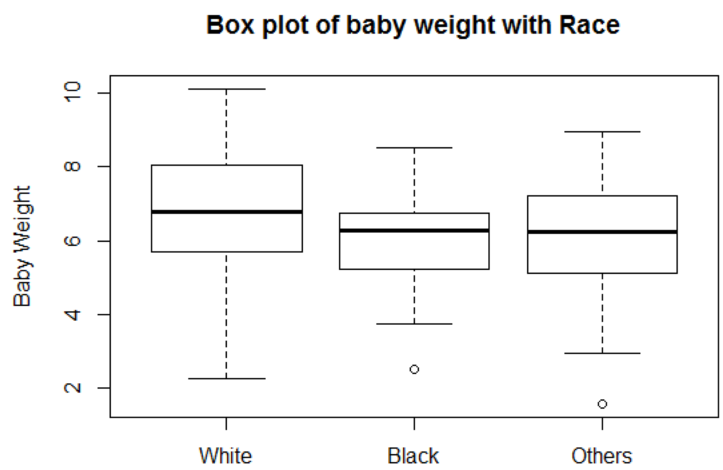
1. **Data Description:**

    a. Weight of the mother:



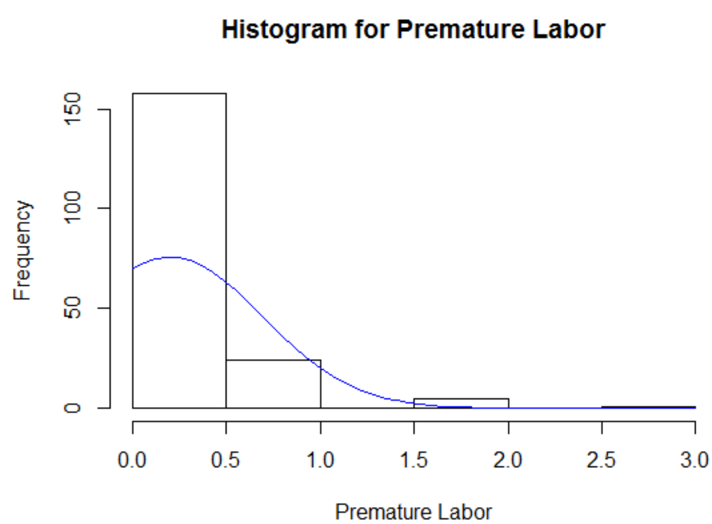
There is linear relationship between baby birth weight and the mother’s weight.

b. Race:



The median baby birth weight of mothers belonging to white race is higher compared to African Americans and others.

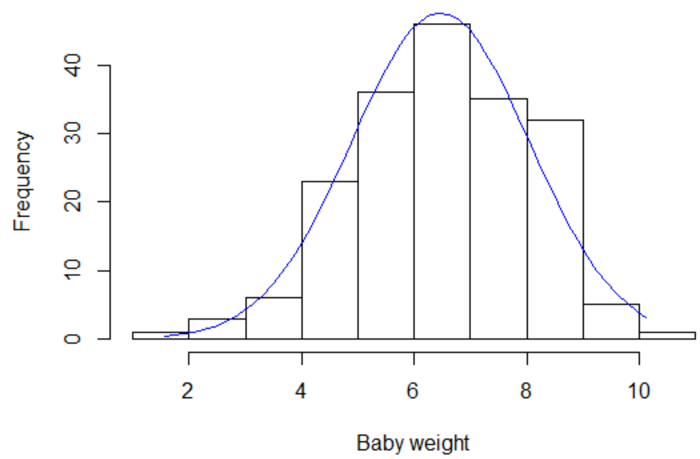
c. Premature Labor:



The distribution of the number of premature births is highly skewed to the right.

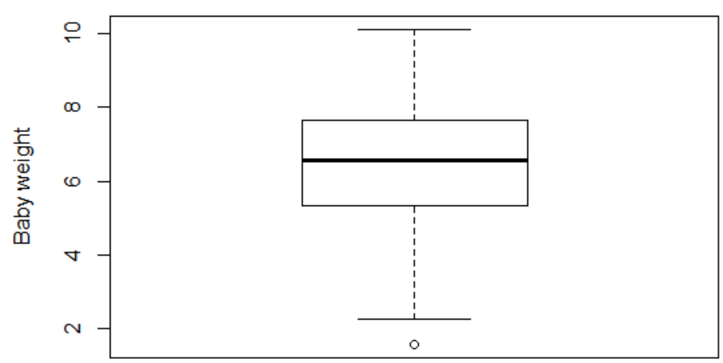
d. Baby birth weight:

Histogram for Baby weight



The distribution of baby birth weights is found to be normally distributed

Box plot of Baby weight



2. Statistical Models:

Model 1:

The first model, say Model 1 includes all the eight factors present in the dataset.

```
##
## Call:
## lm(formula = babywt ~ mother_weight + age + smoke_flag + hyper_flag +
##     uinf_flag + pre_birth + visits + race_factor, data = lowbwt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8845 -1.0037  0.1029  1.0142  3.0704
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.707261   0.682188   9.832  < 2e-16 ***
## mother_weight  0.010558   0.003768   2.802 0.005641 **
## age           -0.026893   0.021840  -1.231 0.219819
## smoke_flag    -0.737121   0.230572  -3.197 0.001644 **
## hyper_flag    -1.310228   0.437461  -2.995 0.003135 **
## uinf_flag     -1.132487   0.300300  -3.771 0.000221 ***
## pre_birth     -0.073513   0.220813  -0.333 0.739587
## visits        -0.015995   0.100614  -0.159 0.873867
## race_factor2  -1.092531   0.324332  -3.369 0.000926 ***
## race_factor3  -0.751555   0.248439  -3.025 0.002853 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.406 on 178 degrees of freedom
## Multiple R-squared:  0.2434, Adjusted R-squared:  0.2052
## F-statistic: 6.364 on 9 and 178 DF,  p-value: 8.278e-08
```

We observe that the probability of getting this model result when there is no relationship between baby birth weight and the above predictors (p-value) is very less. So we can conclude that there is relationship between baby birth weight and at least one of the predictors.
The R^2 is 0.2434 which implies that 24.34% of the variation in baby birth weight is explained by the model. The standard residual error which signifies the average difference between the predicted and observed values of baby birth weight is 1.406 pounds. We have considered the p-value of each predictor to identify their respective significance in explaining baby birth weight. The predictors with p-values which are greater than 0.05 (industry standard) are removed in the next model. We observe from the regression output that age of mother, premature labor history and number of physician visits are not statistically significant.

Model 2:

We remove the above three factors and checked the output of Model 2.

```
##
## Call:
## lm(formula = babywt ~ mother_weight + smoke_flag + hyper_flag +
##     uinf_flag + race_factor, data = lowbwt)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.014 -0.949  0.150  1.015  2.860
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.145756   0.530599  11.583  < 2e-16 ***
## mother_weight  0.009659   0.003638   2.655 0.008642 **
## smoke_flag    -0.731662   0.225375  -3.246 0.001393 **
## hyper_flag    -1.280367   0.433338  -2.955 0.003546 **
## uinf_flag     -1.136622   0.292407  -3.887 0.000142 ***
## race_factor2  -1.007635   0.316429  -3.184 0.001708 **
## race_factor3  -0.712171   0.244953  -2.907 0.004101 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.402 on 181 degrees of freedom
## Multiple R-squared:  0.2352, Adjusted R-squared:  0.2099
## F-statistic: 9.278 on 6 and 181 DF,  p-value: 7.256e-09
```

We observe that the probability of getting this model result when there is no relationship between baby birth weight and the above predictors (p-value) is very less. So we can conclude that there is relationship between baby birth weight and at least one of the predictors.
The R^2 is 0.2352 which implies that 23.52% of the variation in baby birth weight is explained by the model. The standard residual error which signifies that the average difference between the predicted and observed values of baby birth weight is 1.402 pounds. We have considered the p-value of each predictor to identify their respective significance in explaining baby birth weight. We observe that all the predictors in the model are statistically significant.

Model 3:

We then considered only nominal variables, smoking, hypertension, uterine irritability and race and ran an ANOVA (model 3) to check if these factors better explain the baby weight.

```
##                    Df Sum Sq Mean Sq F value   Pr(>F)
## factor(race)        2   21.7  10.839   5.337 0.005591 **
## factor(smoke_flag)  1   31.4  31.416  15.469 0.000119 ***
## factor(uinf_flag)   1   30.9  30.922  15.225 0.000134 ***
## factor(hyper_flag)  1   11.5  11.548   5.686 0.018125 *
## Residuals         182  369.6   2.031
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We observe that the probability of getting this model result when there is no relationship between baby birth weight and the above predictors (p-value) is 0.018 (<0.05). So we can conclude that there is relationship between baby birth weight and at least one of the predictors.

The R^2 is 0.2054 which implies that 20.54% of the variation in baby birth weight is explained by the model. The standard residual error which signifies that the average difference between the predicted and observed values of baby birth weight is 1.4251 pounds. The observed R^2 is less than the previous model and residual error is more than previous model.

Model comparison:

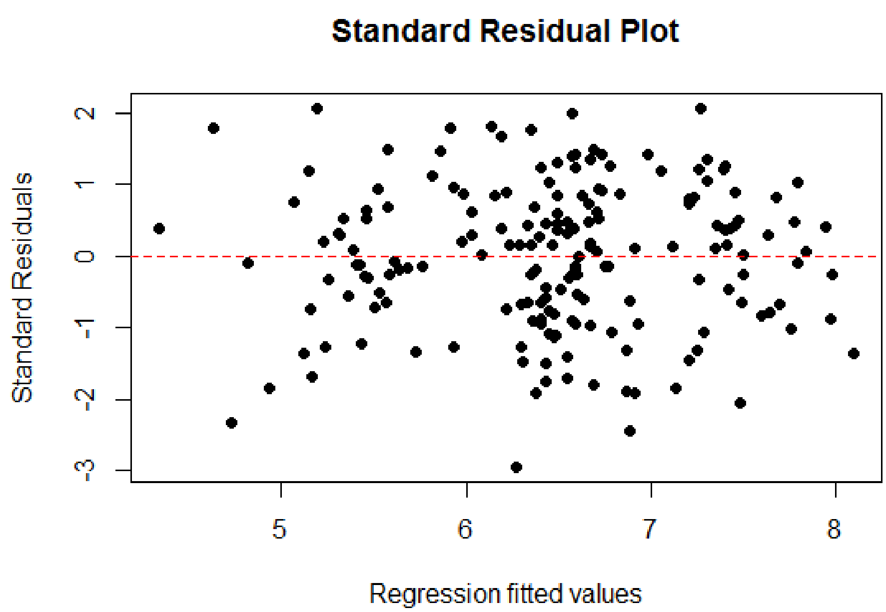
Model	R^2	s	Highest p-value
Model 1	0.2434	1.406	0.874
Model 2	0.2352	1.402	0.009
Model 3	0.2054	1.425	0.018

From the above analysis, we can conclude that model 2 is able to best explain the relationship between factors and baby weight

3. Assumptions for Best Fit Model

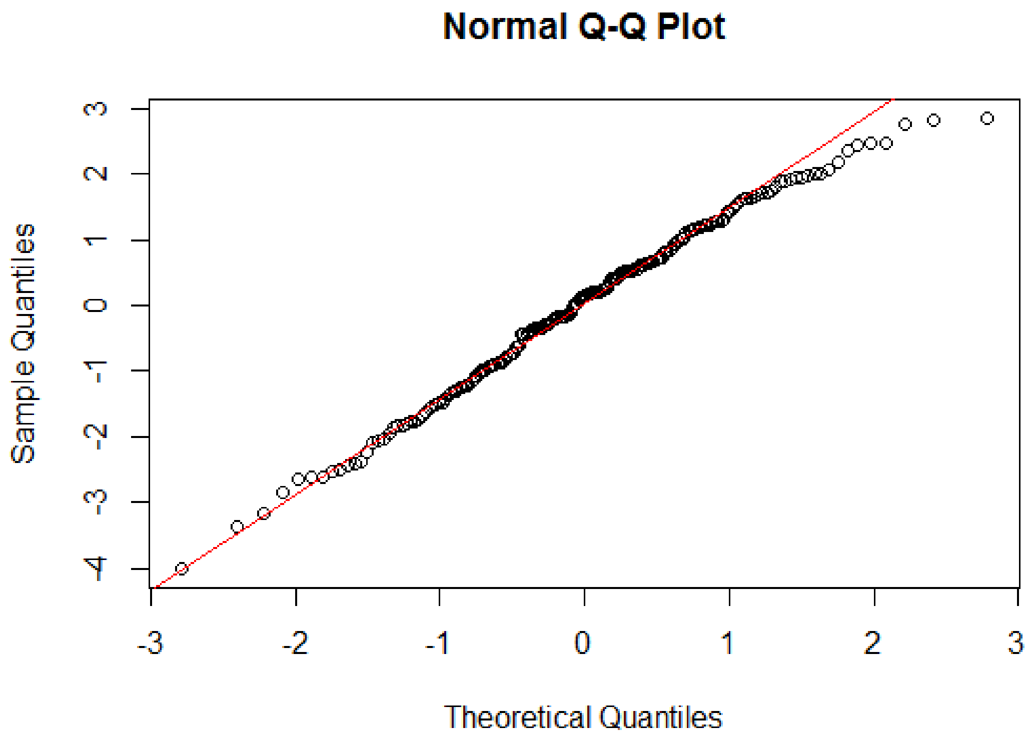
- The data is obtained using random sampling at Baystate Medical Center
- The process is stable over time
- The standard errors are normally distributed with constant standard deviation of error terms over all possible values of the predictors. The normal quantile plots and goodness of fit tests have been performed for all the samples to check for normality.

  Standardized residual plot:



The residuals are equally distributed around zero and there is no particular pattern of the residuals. So the residuals have a mean of zero and standard deviation is constant for all the points.

Normality plot:



From the graph we can conclude that the residuals are normally distributed.

Goodness of fit test:

```
##
##  Shapiro-Wilk normality test
##
## data:  r$residuals
## W = 0.9895, p-value = 0.1824
```

Null Hypothesis: Residuals are normally distributed
Alternate Hypothesis: Residuals are not normally distributed

The p-value is quite high, so we reject the null hypothesis. We can conclude that the residuals are normally distributed.

Collinearity:

We tested the collinearity between the mother's weight and each of the nominal variables. Below is a sample output of the same.

```
## Analysis of Variance Table
##
## Response: mother_weight
##                   Df Sum Sq Mean Sq F value  Pr(>F)
## factor(uinf_flag)   1   4133  4133.0  4.4794 0.03564 *
## Residuals         186 171619   922.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#R Square Calculation
anova(fit)[["Sum Sq"]][1]/(anova(fit)[["Sum Sq"]][1]+anova(fit)[["Sum Sq"]][2])
```

```
## [1] 0.02351626
```

We observe that the R^2 is less than 0.2 which implies there is no collinearity between the mother's weight and uterine irritability. We observed a similar relationship for the other nominal variables as well.

We also tested the collinearity among each of the nominal variables. Below is a sample output of the test between race and smoking status.

Even though we observe that there is a relationship between race and smoking status, we cannot quantify the magnitude of the relationship and hence we need to be cautious about the interpretation of the model.

```
##
##  Pearson's Chi-squared test
##
## data:  table
## X-squared = 22.335, df = 2, p-value = 1.413e-05
```

4. References

1. http://www.umass.edu/statdata/statdata/stat-rmult.html
2. http://ajcn.nutrition.org/content/85/2/584S.long